

## DOCUMENT RESUME

ED 295 801

SE 049 172

AUTHOR Mestre, Jose; And Others  
TITLE Comprehending Premises: Effects of Negations and Training among Anglos and Hispanics.  
SPONS AGENCY National Inst. of Education (ED), Washington, DC.  
PUB DATE 87  
GRANT NIE-83-0072  
NOTE 55p.  
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.  
DESCRIPTORS Bilingual Education; \*Bilingual Students; \*College Mathematics; Concept Formation; Educational Research; Higher Education; \*Hispanic Americans; \*Logic; Logical Thinking; \*Mathematical Logic; Mathematics Education

## ABSTRACT

College students who were either monolingual speakers of English or bilingual Hispanics participated in two studies dealing with the comprehension of sentences containing negations. Both the influence of different sentence constructions on comprehension, as well as the heuristics and misconceptions exhibited by subjects as they attempted to comprehend the sentences were investigated. The second experiment was designed primarily to determine whether it was possible to develop an intervention strategy that eliminated the use of erroneous strategies. In the first experiment, bilinguals took longer than monolinguals to comprehend sentences containing negations, although there were no interactions between sentence construction and language background that would suggest the use of qualitatively different strategies by the two groups. The intervention procedure used in the second experiment was extremely effective for monolinguals and for most bilinguals, as measured by performance a week later. When subjects returned to be tested six months later, the improvement in performance was maintained for the monolinguals, but not the bilinguals. The educational implications of these findings are discussed. (Author)

\*\*\*\*\*  
\* Reproductions supplied by EDRS are the best that can be made \*  
\* from the original document. \*  
\*\*\*\*\*

ED 295801

COMPREHENDING PREMISES: EFFECTS OF NEGATIONS AND  
TRAINING AMONG ANGLOS AND HISPANICS

Jose Mestre, Pamela Thibodeau Hardiman,  
William Gerace and Arnold Well

Department of Physics & Astronomy  
University of Massachusetts  
Amherst, MA 01003

U.S. DEPARTMENT OF EDUCATION  
Office of Educational Research and Improvement  
EDUCATIONAL RESOURCES INFORMATION  
CENTER (ERIC)

☒ This document has been reproduced as  
received from the person or organization  
originating it.  
☐ Minor changes have been made to improve  
reproduction quality.

• Points of view or opinions stated in this docu-  
ment do not necessarily represent official  
OERI position or policy.

"PERMISSION TO REPRODUCE THIS  
MATERIAL HAS BEEN GRANTED BY

*Jose*  
*J. Mestre*

TO THE EDUCATIONAL RESOURCES  
INFORMATION CENTER (ERIC)."

ABSTRACT

College students who were either monolingual speakers of English or bilingual Hispanics participated in two studies dealing with the comprehension of sentences containing negations. Both the influence of different sentence constructions on comprehension, as well as the heuristics and misconceptions exhibited by subjects as they attempted to comprehend the sentences were investigated. The second experiment was designed primarily to determine whether it was possible to develop an intervention strategy that eliminated the use of erroneous strategies. In the first experiment, bilinguals took longer than monolinguals to comprehend sentences containing negations, although there were no interactions between sentence construction and language background that would suggest the use of qualitatively different strategies by the two groups. The intervention procedure used in the second experiment was extremely effective for monolinguals and for most bilinguals, as measured by performance a week later. When subjects returned to be tested six months later, the improvement in performance was maintained for the monolinguals, but not the bilinguals. The educational implications of these findings are discussed.

BEST COPY AVAILABLE

SE 647 172

Reasoning according to the rules of logic does not come naturally to many people, as numerous studies of syllogistic reasoning have shown (Chapman and Chapman, 1959; Johnson-Laird, 1983; Johnson-Laird and Steedman, 1978; Wason and Johnson-Laird, 1972; Woodworth and Sells, 1935). There is evidence that subjects employ a variety of heuristics in reasoning tasks, often involving misconceptions. Certain populations, such as Hispanic bilinguals, are particularly prone to using such heuristics (Mestre, 1982, 1986). A related finding is that the appearance of negatives in syllogisms has an adverse effect on performance (Johnson-Laird and Steedman, 1978). Negations are also known to adversely affect sentence comprehension (Sherman, 1973, 1976). Given the relative importance of logical reasoning skills and research evidence indicating a number of difficulties experienced by students when dealing with negations, the present study was designed to: 1) systematically study the strategies and misconceptions exhibited by monolingual and Hispanic bilingual college students in comprehending premises containing negations, and 2) determine whether it is possible to design an intervention strategy that eliminates the use of erroneous strategies.

A number of research studies indicate that the presence of negations imposes increased memory loads in the processing of sentences. For example, sentences containing negations generally take longer to verify than affirmative sentences (Just and Carpenter, 1971; Slobin, 1965; Trabasso, Rollins and Shaughnessy, 1971; Wason, 1959, 1961). The single exception to this result occurs in single negative, context-specific instances, called "plausible denials," in which the negative statements took less time to verify than comparable affirmative statements (Wason, 1965; Arroyo, 1982). One explanation for the increased latencies in comprehending negations is that the

processing of negations requires a transformation that is additional to the processing of affirmatives (Gough, 1965; Miller, 1962; Miller & McKean, 1964).

Sentences containing negations are also harder to recall (Mehler, 1963). In accord with the hypothesis that the processing of negations requires a transformation, Clark and Card (1969) report that subjects found sentences easier to remember if they transformed negations into affirmations. Subjects also convert double negations containing prefixes such as in- and un- into affirmatives, for example, changing "not unprepared" to "prepared" (Sherman; 1973, 1976). Vazquez (1981) found a similar cancelling strategy; subjects cancelled the two negations in a statement such as "it isn't true that the dots aren't red." However, Vazquez also found that subjects were not predisposed to cancel the two negatives when "false" replaced "not true," as in, "it's false that the dots aren't red."

Research findings also indicate that increasing the number of negations results in successive decrements in comprehension (Johnson-Laird, 1970; Legrenzi, 1970, Sherman, 1973, 1976), with sentences containing three or more negations beyond the normal comprehension ability of subjects. The inability to comprehend more than two negations has been described by Sherman (1976) as "cognitive overload." Sherman claims that combining two negatives to form an affirmative takes effort and space in the mental sentence processor and that when a third negative is encountered, a system overload results and the "two-negative = affirmative" coding cannot be maintained.

The explanations discussed above that have been offered to account for the increased latencies and difficulties in dealing with negations are consistent with the assumption that the human reasoning system is rational, but limited in working memory capacity (see Johnson-Laird, 1983 for a fuller explication). Many of the erroneous strategies and misconceptions that have

been uncovered in syllogistic reasoning research are also consistent with the rational, but memory-limited processor model. One example from syllogistic reasoning is the "atmosphere hypotheses" (Woodworth and Sells, 1935), where subjects attempt to match the quantifiers of the premises to those of the conclusions. Another erroneous strategy, called the "figural effect" (Johnson-Laird and Steedman, 1978), proposes that subjects prefer those conclusions that can be constructed by forming a continuous, transitive link from the premises, similar to the mathematical rule which allows us to conclude that  $a > c$  if  $a > b$  and  $b > c$ . Therefore, strategies such as quantifier matching and transitive linking, have evolved to overcome these limitations in short term memory.

In fact, subjects' habitual use of the "rules" governing the comprehension of natural discourse in reasoning tasks could also be interpreted as an attempt to reduce memory load. Braine (1978) claims that common fallacies in syllogistic reasoning are due to introducing habits characteristic of practical reasoning and natural language comprehension into formal reasoning. Henle (1962) observed that subjects tend to prefer "empirically reasonable propositions" to "logical" ones. In accord with the view that limits in working memory constrain reasoning, Henle claimed that errors in reasoning may not reflect an inability to reason, but rather a tendency to: 1) simplify situations by disregarding premises, 2) interpret premises in an unintended way, 3) disregard the logical task, or 4) introduce outside knowledge as an additional premise.

Previous research in the comprehension of premises indicates that many college students, particularly Hispanic bilinguals, have inordinate difficulty comprehending premises with negations (Mestre, 1982, 1986). The focus of the present set of experiments was two-fold. First, we wished to systematically

investigate monolingual and Hispanic bilingual undergraduates' ability to comprehend premise-like statements as a function of the number and type of negation embedded within the statement. Second, after determining the types of misconceptions exhibited by students and the frequency of their use, we were interested in ascertaining whether such misconceptions could be eliminated through a short intervention strategy.

There were several practical reasons for choosing each of these two areas for investigation. Insofar as both reasoning and comprehension of negations cause difficulties for the general populace, it is important to investigate ability to comprehend single premises containing different patterns of negations. Identifying common erroneous strategies and misconceptions exhibited by subjects in interpreting premises is a first step toward understanding the processes and erroneous strategies in more complex tasks, such as syllogistic reasoning. After identifying the causes of difficulty, the problem evolves into one of instructional design. In particular, we were interested in two related questions: Can we utilize our knowledge of subjects' difficulties and misconceptions in designing an appropriate remedy that proves effective in the long-run?, and if the answer to this last question is "yes," What are the factors that help make this remedy effective? For Hispanics, the task of identifying intervention strategies that improve general cognitive performance is important since language deficiencies often militate against their success in college level technical fields of study where ability to reason is imperative (Mestre, 1981, 1986).

Studying the performance of Hispanics on comprehension tasks containing negations could prove interesting for a related reason: the English parallels of certain Spanish double-negative constructions contain only a single

negation.<sup>1</sup> In Spanish, these doubly negated constructions retain a negative meaning instead of reverting to an affirmative meaning, as would be the case in grammatically correct English. For example, the Spanish translation of "I do not want any money" is "yo no quiero ningún dinero." Translated literally into English, this Spanish statement would become "I do not want no money." The Spanish negations "no" and "ningún" (meaning "no" and "none", respectively) result in an overall negative meaning when they appear together in a sentence. However, these negations also yield a negative meaning when they appear alone.

In Spanish, it is also possible to add negations without altering the meaning of sentences, as the following example will illustrate. The two-negative construction "neither/nor" has a completely equivalent and parallel counterpart in Spanish, namely "no/ni." The translation of a statement such as "I want neither money nor fame" into Spanish becomes "yo no quiero dinero ni fama." However, it would be perfectly acceptable in Spanish to include an additional negation in this statement without changing its meaning by writing "yo no quiero ni dinero ni fama." To do so in English would result in the awkward and confusing statement "I do not want neither money nor fame" meaning (in grammatically correct English) "I want both money and fame."

Differences in the structure of languages have been observed to influence comprehension for certain tasks. Three-, and four-year old native speakers of English and Japanese performed differently when asked to verify examples of four types of statements (Akiyama: 1984, in press): 1) true affirmatives ("you are a child"), 2) false affirmatives ("you are a baby"), 3) false negatives

---

<sup>1</sup> This should not be taken to mean that a complete lack of parallelism exists. Only in certain cases, such as those illustrated above, do English and Spanish differ in the meaning of double negations.

("you aren't a child"), and 4) true negatives ("you aren't a baby"). The pattern of performance by English speaking children paralleled that observed with adults (Carpenter and Just, 1975): true negatives were the most difficult to verify, followed by false negatives, false affirmatives, and true affirmatives. In contrast, Japanese children found false negatives most difficult to verify, followed by true negatives, false affirmatives and true affirmatives. Akiyama attributed this difference in performance to specific differences related to negations between English and Japanese. In the present study we were interested in investigating whether the aforementioned differences between English and Spanish give rise to a distinctly different pattern of performance between Anglos and bilingual Hispanics in double negative tasks phrased in English. Given that the use of double negatives in the English language is not uncommon, any differences in performance between Anglos and Hispanics would have important pedagogical ramifications.

#### EXPERIMENT 1

Stated in general terms, the focus of Experiment 1 consisted of ascertaining the extent to which performance on comprehension tasks depended both on the number and type of negation, and on linguistic factors, such as sentence context and phrasing. The subjects read relatively short stimulus sentences containing one, two and three negations, and then selected the one of four positively phrased multiple choice answers that correctly paraphrased the original sentence.

Three types of negations appeared in the sentences: 1) falsification, or superordinate negation, in which the entire sentence was negated, as in "it's not true that ... .", 2) subject negation, or quantifier negation, as in "not all ... .", and 3) predicate or category negation as in "... are not male." Eight different sentence types were constructed with various possible

combinations of these three types of negations. An example of a sentence containing both falsification and subject negations is: "It's not true that not all clerks working at the Fitzgerald Company are male."

Manipulation of number of negations will allow us to determine the relative difficulty of one-, two-, and three-negation sentences, as well as to ascertain whether there is a difference between the performance patterns of monolinguals and bilinguals. If present, a negation by group interaction would support the hypothesis that a lack of parallelism between Spanish and English results in comprehension differences. A group main effect would be consistent with the findings of Hunt (1978), who observed that negative wording has a stronger adverse affect on comprehension for low verbal subjects than for high verbal subjects. Previous research with the population of bilingual Hispanic engineering majors in the same university indicates that these students are relatively underprepared in language skills in comparison to their monolingual counterparts (Mestre, 1981).

The second linguistic factor investigated in Experiment 1 was the phrasing of the stimulus sentences. Three different phrasings were used. The standard phrasing integrated the context into the body of the sentence, as in "It's not true that not all of the clerks working at the Fitzgerald Company are male." It is possible that separating the context from the body of the sentence might lessen the load on working memory as well as focus attention on the sentence's deep structure; if this is the case, the result would be an improved performance over the phrasing where the sentence's context and deep structure are integrated. Thus, the context-separated phrasing used was "Concerning the clerks working at the Fitzgerald Company, it's not true that not all are male." A third type of phrasing was used to investigate whether changing the type of negation influences the propensity to use a two-negative

cancelling strategy. Recall that the phrasing, "it isn't true that all the dots aren't red," has been observed to elicit a cancel-two-negatives strategy, whereas, "it's false that all the dots aren't red," did not (Vasquez, 1981). Therefore, the third phrasing was "It's false that not all of the clerks working at the Fitzgerald Company are male."

The third linguistic factor investigated was the type of category used within the context of the sentence. Two types of categories were used, and both dichotomized the problem space into two distinct partitions. One type of category consisted of single words, such as the gender categories male/female and glass-type categories tinted/clear; we call this type of category "discrete," since there are always only two values that they can possibly possess. The other type of category consists of multi-word phrases, such as "get more than twenty miles per gallon" referring to car mileage; we call this type of category "continuous" since there is a continuous range of allowable values for car mileage. It was expected that the discrete categories would impose a lighter memory load, and therefore prove easier than the continuous categories.

#### Method

##### Subjects

Seventy-one undergraduate science and engineering majors from the University of Massachusetts participated in Experiment 1. Each subject received \$5.00 at the end of the experimental session.

The subjects were divided into five experimental groups. Group B1 consisted of 11 bilingual Hispanic students who were conversantly fluent in both English and Spanish and received stimulus sentence set 1. Their mean grade point average (G.P.A.) was 2.94 (s.d. .27). The remaining 60 subjects

were monolingual speakers of English. Their mean G.P.A. was 3.06 (s.d. .47). The monolingual subjects were divided into four equal-sized groups, M1, M2, M3, and M4, and were given stimulus sentence sets 1, 2, 3, and 4, respectively.

### Materials

The subjects were presented with 32 sentences of similar structures, which contained either one, two, or three negations. They were to indicate which of a set four alternatives had a meaning most similar to the target sentence. The sentences were presented in one of four forms, Stimulus Sentence Set 1, 2, 3, and 4. Each subject saw sentences of only one form.

Stimulus Sentence Set 1 The target and alternative sentences for Stimulus Sentence Set 1 were structured as follows (round brackets indicate optional portions of the sentence):

(Falsification) (subject negation) [quantifier] [group of objects]  
[verb] (predicate negation) [dichotomous category]

The sentence parts varied as follows:

Falsification - "It's not true that" (used in sets 1, 2, and 4) and "It's false that" (set 3)

Subject and Predicate Negations - "not" (sets 1, 2, and 4) or contraction with verb, such as "aren't" (set 3)

Quantifier - "all" or "some"

Group of Objects, Verb, and Dichotomous Category - These three portions of the sentence together formed the Context of the sentence. The four Contexts were:

1. the water samples / contain / more than three types of bacteria
2. the cars / get / more than twenty miles per gallon
3. clerks working at the Fitzgerald Company / are / male
4. windows in the Martin Tower / are / tinted

The complements of the dichotomous categories for each Context were:

1. three or fewer types of bacteria
2. twenty or fewer miles per gallon
3. female
4. clear

Note that two types of dichotomous category labels were used : 1) the Continuous categories used in Contexts 1 and 2 consisted of a range of possible values that had an arbitrary split point, and 2) the Discrete categories used in Contexts 3 and 4 had only two values that differed qualitatively.

Twelve possible Sentence Types can be constructed using this general form, if the awkward construction "not some" is avoided. Eight of the twelve Sentence Types were used as stimulus sentences, while the remaining four Sentence Types comprised the multiple choice answer set. The Sentence Types for the multiple choice answer set were modified slightly, such that for the two cases in which a negation preceded the dichotomous category, the category complement was used. Each of the eight Sentence Types appeared in each of the four Contexts, with the appropriate multiple choice alternative set, for a total of 32 problems. Examples of each of the eight Sentence Types and four multiple choice alternatives are presented in Table 1.

-----  
INSERT TABLE 1 ABOUT HERE  
-----

Each of the four multiple choice alternatives was the appropriate answer for two Sentence Types. Table 1 indicates the correct pairing of Sentence Types and multiple choice alternatives.

Stimulus Sentence Sets 2, 3, and 4      The eight Sentence Types and four multiple choice alternatives were modified to produce Stimulus Sets 2, 3, and 4. For Stimulus Set 2, the group of objects was taken out of the body of the sentence and placed at the beginning of both the target sentence and the set of multiple choice alternatives. For example:

Concerning the clerks working at the Fitzgerald Company, it's not true that all are not male.

Concerning the clerks working at the Fitzgerald Company,

- A) All are male.
- B) Some are male.
- C) All are female.
- D) Some are female.

Stimulus Set 3 was identical to Stimulus Set 1 except that the falsification, "It's not true that" was replaced by "It's false that," while the verbal negation, "are not" was replaced by "aren't." For example:

It's false that all of the clerks working at the Fitzgerald Company aren't male.

Stimulus Sentence Set 4 differed from Stimulus Sentence Set 1 by the addition of a fifth multiple choice alternative. Subjects in a previous study (Mestre, 1986) indicated a "some of each" alternative would be useful. This fifth alternative took the form:

- E) Some of the clerks working at the Fitzgerald Company are male, and some are female.

### Procedure

The experimental session was approximately 30 to 45 minutes long. Subjects were run on a Wang VS-80 minicomputer; two subjects participated in the experiment at the same time. All bilingual subjects received Stimulus Sentence Set 1, while the 60 monolingual subjects were randomly assigned to groups receiving Stimulus Sets 1, 2, 3, and 4. These groups are referred to as M1, M2, M3, and M4, respectively.

The instructions were written and included the following sample problem: "Some student solutions to the quiz were not written in ink." The subjects were instructed to pick the multiple choice alternative that conveyed "the equivalent information as the alternative," and respond by typing the letter of the choice into the computer. The instructions familiarized the subjects with both the contexts and the multiple choice alternative sets that would actually be used in the experiment. They stressed that although items dealing with the same topic might sound similar, "all questions are slightly different."

The subjects were provided with scrap paper to do any writing that they wished. The time to respond from the presentation of the item on the screen was recorded, but subjects were not informed that their reactions were being timed, in order to assess processing time under relatively natural circumstances. All subjects were given a short break after responding to the first half of the questions.

### Item Presentation

The items were presented randomly within two constraints: 1) Discrete and Continuous Context items were alternated, such that an item from Context 1 or 2 was followed by an item from Context 3 or 4, and 2) the eight Discrete

and eight Continuous items given in each half of the experiment formed a complete set of the eight Sentence Types, i.e. the sets were formed from a mixture of Contexts 1 and 2, and 3 and 4.

After the subject's response to each item, the subject was told: 1) the number of items that had been answered, 2) the number of items remaining, and 3) to press "Return" to see a new item.

## Results

### Preliminary Analysis

In order to simplify later analysis, it was desirable to combine the two Continuous Contexts, 1 and 2, and the two Discrete Contexts, 3 and 4. Comparisons within each group revealed no significant differences between Contexts 1 and 2 or 3 and 4 on the percent correct measure. However, there was a difference between the two Continuous Contexts on the latency measure: the mean response time to the Context 1, or bacteria problems was 47.15 seconds, while the mean response time to Context 2, or gas mileage problems was 42.57 seconds. Although this difference was significant,  $F(1,52)=6.28$ ,  $p<.05$ , the difference was not sizeable. Thus, the two Continuous contexts were combined despite the difference in latency.

Preliminary analysis of the groups also revealed a considerable disparity between the performance of group M4, the group given five multiple choice alternatives, and the other four groups. Group M4 averaged only 13.3 (of 32) correct answers, while the other groups averaged 18.9 correct answers. This difference is significant,  $t(66) = 2.808$ ,  $p < 0.01$ . Obviously, the "some of each" alternative proved to be quite compelling. Since the performance of group M4 was so different from the other groups, this group was not included in the major analysis.

### Group and Stimulus Set

Stimulus Set (M1, M2, M3) x Sentence Type (1-8) x Context (Discrete, Continuous) ANOVA's on percent correct and latency scores for the three M Stimulus Set groups revealed no main effects of Stimulus Set. Pairwise comparisons ( $k=3$ ,  $p=0.017$ ) revealed no significant differences among the groups. Thus, variation in the format of the sentence presentation does not appear to significantly influence either latency or accuracy.

The mean percent correct for the bilingual Hispanic group, B1, on the 32 items, 17.4 (s.d. 6.6), was slightly lower than the average of the three monolingual groups, 19.4 (s.d. 6.2). Although the difference in percent correct was not significant,  $t(52) = 0.900$ , the B1 group had a significantly longer mean latency than the combined M groups, 52.55 versus 37.09 seconds  $t = 1.822$ ,  $p < .05$ .

### Sentence Type

Although the factor of Stimulus Set did not influence performance, the Type of Sentence did have an effect,  $F(7,364) = 9.31$ ,  $p < .001$  (see Figure 1). Overall, the mean percent correct was highest on those Sentence Types with one negation, 64.7, followed by two-negation Sentence Types, 59.5, and the three-negation Sentence Type, 42.4. Sentence Types with one (Types 1, 3, and 7), two (2, 4, 5, and 8), and three negations (6) were combined and analyzed in a Group x Number of Negations x Context ANOVA. There was a main effect of number of negations,  $F(2,104) = 12.74$ ,  $p < .0001$ . The percentage of correct responses was greater for sentences with one and two negations than those with three negations,  $t(70) = 3.72$ ,  $p(\text{Bonferroni}, k=3) = .0012$ , and  $t(70) = 3.34$ ,  $p(\text{Bonferroni}, k=3) = .0042$ , respectively. However, two negations were not significantly more difficult than one negation (See Figure 2). Note by examining Figure 1 an inconsistency in the percent correct data: the two-

negation Sentence Type 5, "It's not true that not all...", had a higher rate of correct responses than two of the three single-negation Sentence Types (i.e. 1 and 3).

-----  
INSERT FIGURE 1 ABOUT HERE  
-----

-----  
INSERT FIGURE 2 ABOUT HERE  
-----

The latency data yielded results that were similar to the percent correct results; the mean latency for one negation items was 32.46 seconds, for two negation items it was 41.48, and for three negation items it was 57.69. An examination of Figure 3 indicates that these data did not violate the tendency of difficulty to increase with number of negations. Each of the mean response times to Sentence Types with one negation was shorter than each mean response for Sentence Types with two negations. Thus, additional negations add to the processing time.

-----  
INSERT FIGURE 3 ABOUT HERE  
-----

### Context

The Context of the sentence, i.e. Discrete or Continuous, also appeared to play an important role in sentence understanding. The mean percent correct on the Discrete items, 62.6, was significantly higher than the mean percent correct on the Continuous items, 55.4,  $F(1,52) = 12.41$ ,  $p < .001$ . For six of the eight Sentence Types, performance was better on the Discrete items (See Figure 1). In fact, the only exception was Sentence Type 4, for which performance was actually significantly better in the Continuous condition,  $t(70) = 3.05$ ,  $p = .0256$  (Note:  $p$  values are for a Bonferroni planned comparison test with  $k=8$ ). Performance was significantly better in the Discrete condition for Sentence Types 3,  $t(70) = -3.34$ ,  $p = .0104$ , 7,  $t(70) = -5.03$ ,  $p < .0001$ , and 8,  $t(70) = -4.11$ ,  $p = .001$ .

The influence of Context on the latency data was similar, but reveals a clearer pattern: the mean latency was longer for Continuous than Discrete contexts, 45.8 seconds versus 36.1 seconds,  $F(1,52) = 56.54$ ,  $p < .0001$ . The mean latency was shorter in the Discrete Context for all Sentence Types (See Figure 3). The difference between Contexts was significant for all Sentence Types except 2 and 6 ( $p$  values are for Bonferroni comparisons with  $k=8$ ): 1 -  $t(70) = 3.26$ ,  $p = .0136$ , 3 -  $t(70) = 2.80$ ,  $p = .05$ , 4 -  $t(70) = 3.74$ ,  $p = .0032$ , 5 -  $t(70) = 3.22$ ,  $p = .0152$ , 7 -  $t(70) = 3.98$ ,  $p = .0016$ , and 8 -  $t(70) = 5.43$ ,  $p < .0001$ .

Context also interacted with number of Negations on the percent correct measure,  $F(2,104) = 5.85$ ,  $p = .0039$  (See Figure 2). Performance was significantly better in the Discrete context for sentence types with one negation  $t(70) = -6.07$ ,  $p$  Bonferroni,  $k=3$   $< .0001$ , but did not vary as a function of Context for sentences with two or three negations.

### Practice Effects

Percent correct on the first and second halves of the experiment were compared to determine whether there was an effect of practice. The mean percent correct on the second half, 62.2%, was significantly better than in the first half, 56.5%,  $F(1,52) = 6.74$ ,  $p=0.0122$ , although the actual level of improvement was not remarkable.

Subjects also became faster during the second half of the experiment: the mean response latency was 48.2 seconds on the first half and 32.2 seconds on the second half. There was a Group x Time interaction,  $F(3,52) = 14.60$ ,  $p<.001$ , indicating a difference in how much faster each group became: 9.6, 12.2, 20.4, and 26.4 seconds faster, for groups M1, M2, M3, and B1, respectively.

### Response Preference

The request of subjects from previous studies for a "some of each" multiple choice response suggested that subjects may prefer alternatives which seem less committal. Accordingly, "some" versus "all" responses were analyzed in a Response Preference x Group x Context ANOVA. Subjects displayed a significant preference for "some" answers, i.e. choices B and D,  $F(1,52)=5.08$ ,  $p<.05$ . An average of 17.09 of the 32 responses were "some" responses, while 14.91 were "all" responses. Context also interacted with Response Preference,  $F(1,52)$ ,  $p=.0017$ , such that subjects gave more "some" responses for Continuous context items than Discrete context items.

### Discussion

The present study indicates that negations play an important role in the ease of comprehension. Consistent with previous studies (Just & Carpenter,

1971; Slobin, 1965; Sherman, 1973, 1976; Trabasso et al., 1971; Wason, 1959, 1961), the length of time needed to interpret a sentence increased significantly as the number of negations increased: Sentences containing three negations required approximately 15 seconds more than those containing two negations, which in turn required approximately 10 seconds more than one-negation sentences. This trend in the latency data suggests that difficulty increases as the number of negations increases. However, the percent correct data do not unambiguously support this claim. Although accuracy was significantly lower for the three-negation sentences than for either the one- or two-negation sentences, one- and two- negation sentences did not differ significantly from one another. This finding is quite surprising in view of latency data from the present experiment, as well as previous research findings (Johnson-Laird, 1970; Legrenzi, 1970; Sherman, 1973, 1976).

This contradictory result appears to be largely a result of the strikingly high accuracy on the Sentence Type 5, which had two negations of the form "It's not true that not all ..." (See Figure 1). The mean percent correct for items of this type was 75%, the best performance among the eight sentence types. Such high performance may be attributable to two factors: 1) the "cancel-two-negations" strategy used to make a positive statement (Mestre, 1986; Sherman, 1973, 1976; Vazquez, 1981) which yields a correct answer only for Sentence Type 5, and 2) a "prompting effect," by which the close proximity of two negations invites subjects to use the cancel-two-negations strategy in Sentence Type 5 and not in other sentence types. Given that the subjects were science and engineering majors and quite facile with mathematics, this prompting influence may be an obvious strategy, since in mathematics two juxtaposed negatives do result in a positive (e.g.  $-(-5)=+5$ ). Together,

these two factors would produce an abnormally high number of correct responses in Sentence Type 5.

Despite evidence that subjects were predisposed to cancel negations in pairs under certain circumstances, there was no evidence that subjects used different strategies in processing the two forms "it's not true that..." and "it's false that...". The research of Vazquez (1981) would suggest that subjects might use a "cancel-two-negatives" strategy in "It's not true that not all.." statements, but not in "It's false that not all..." statements. However, the performance pattern was not influenced by substituting "false" for "not true."

Experiment 1 offered no indication that bilingual Hispanics and monolingual English speakers process double negations differently. The difference could have been absent for two reasons. It is possible that bilingual Hispanics and monolingual English speakers do process negations differently under some conditions, but that the stimulus sentences used in Experiment 1 failed to elicit a double-negation interference effect in Hispanics. Alternatively, it is possible that bilingual Hispanics who have reached the age and attained the English proficiency of the subjects in this study have sorted out nuances in the meaning of negations within the three logical systems they frequently use: English, Spanish and mathematics. When differences in processing negations have been observed, the subjects have been very young (three-, and four-year olds) (Akiyama: 1984, in press). It is less likely that such subjects would be cognizant of nuances in meaning of negations.

Performance on this task was clearly much better than chance (25% correct), however, it was far from optimal despite improvements during the second half of the experiment. Manipulation of surface structure features by

v  
varying sentence phrasing had little effect on performance, indicating that sentence difficulty is determined by the deep structure of the sentence (i.e. its meaning kernel), rather than by the actual wording. The fact that this attempt to focus attention on the critical portion of the sentence did not improve performance suggests that a sentence's meaning kernel is more responsible for imposing memory loads which affect processing than are the sentence's context or phrasing.

Experiment 1 also revealed that Context has an influence on comprehension; sentences with Discrete categories were processed faster and with fewer errors than sentences with Continuous categories. These differences may be related to both the number of words comprising the category and the relative difficulty in transforming "not Category A" to the positive, category complement. Discrete categories, having single word descriptors, may be easier to remember because they are shorter. The complementary categories are also single words, making transformations involving negations substantially easier to implement. For example, the transformation necessary to turn the meaning kernel, "not true, all clerks are male," into an affirmative kernel is relatively simple. If "all clerks are male" is not true, then there must be at least one clerk who is female, yielding "some clerks are female" as the desired answer. In contrast, if it is not true that "all water samples contain more than three types of bacteria," then at least one water sample must be a member of the category that is opposite to "more than three types of bacteria." Determining the category complement requires a more complicated logical transformation, i.e. "more than" must be transformed to "less than or equal to."

The data strongly suggest that subjects use a variety of strategies to paraphrase negative sentences into the affirmative. One factor that

influences choice of strategy is the sentence's surface structure. The extremely good performance on Sentence Type 5 (it's not true that not all ...) suggests that the proximity of the two negatives serves as a prompt for subjects to select a cancel-two-negatives strategy. A second factor influencing strategy selection consists of the use of the rules governing natural discourse comprehension in paraphrasing sentences. For example, in Sentence Type 1 (e.g. Not all clerks ... are male), nearly all incorrect answers consisted of the selection "Some clerks ... are male" suggesting the natural discourse strategy in which "not all" is simply replaced by "some." In summary, the findings suggest that subjects were not as facile in processing the sentences used in the task as their performance might indicate.

## EXPERIMENT 2

Experiment 1 revealed that subjects are not consistently successful in comprehending premises containing negations. They appear to use such strategies as inappropriately applied rules of logic (e.g. cancelling two negations), or natural discourse logic (e.g. replacing "not all" with "some"). In some cases surface structure appears to cue strategy use. This type of problem solving behavior is typical of novices and has been observed in several domains such as physics (Chi, Feltovich and Glaser, 1981) and mathematics (Schoenfeld and Herrmann, 1982).

Experiment 2 was designed to determine whether an intervention strategy could be developed incorporating the findings of Experiment 1 and previous research (i.e. Anzai and Simon, 1979; Mestre, 1986) that would allow subjects to perform well on comprehension tasks similar to those of Experiment 1. The intervention strategy that was developed involved teaching subjects a right-to-left procedure for parsing the sentences. Think-aloud interview data from

a previous study (Mestre, 1986) revealed that subjects who consistently obtained correct answers used a right-to-left parsing procedure. Subjects were also given an opportunity to practice the parsing procedure during the course of the intervention strategy.

The intervention addressed explicitly the incorrect strategy of converting "not all" to "some." The common incorrect strategy is to merely change "not all" to "some" without changing the category to its complement. Further, the intervention attempted to focus the subjects' attention on the sentences' meaning through the use of concrete examples and rough diagrams (not Venn diagrams). This intervention was implemented in a half-hour videotaped lesson.

The real test of the effectiveness of an intervention strategy is the pervasiveness of the effects over time. If a short intervention strategy could be shown to have a lasting effect on performance, it would be of considerable pedagogical importance. For this reason, subjects were tested at one week and again at six months following the viewing of the lesson. Therefore, subjects participated in either three or four experimental sessions: Session 1 was used to administer the 32 sentence task using a paper-and-pencil format; Session 2 was used to present the video lesson and readminister a different randomization of the 32 sentence task; Session 3, one week following Session 1, was used to administer the 32 sentences a third time; Session 4, six months following the lesson, was used to administer the 32 sentences one last time.

#### Method

##### Subjects

Fifty-seven undergraduate science and engineering majors from the University of Massachusetts participated in the first three sessions of

Experiment 2. Twenty-five students returned to participate in a fourth session six months later. Each subject received \$5.00 for participating in each of the sessions, for a total of \$15.00 or \$20.00.

The subjects were divided into three experimental groups: 1) Group BE - 17 bilingual Hispanic students who saw the videotaped lesson, 2) Group ME - 20 monolingual students who saw the videotaped lesson, and 3) Group MC - 20 monolingual students who did not see the videotaped lesson. Groups ME and MC were formed by selecting pairs of subjects who had performed approximately equally well on Session I, and randomly assigning one member of the pair to one of the two groups. The mean quantitative and verbal SAT scores, as well as the G.P.A.'s for each group are presented in Table 2.

-----  
INSERT Table 2 ABOUT HERE  
-----

### Materials

In each of the experimental sessions, subjects were given the same set of 32 written sentences in a booklet. Two randomized orders of the items were used: Randomization A was used in Session I, B in Session II, A in Session III, and A or B in Session IV. These items were nearly identical to those used in Experiment 1; the Continuous context items were slightly modified so that they would be more parallel in form to the Discrete items. The modified Continuous contexts were:

- 1) water samples taken from Sunset Lake/ contain/ more than three  
types of bacteria
- 2) cars in the ACE Parking Lot/ get/ more than twenty miles per  
gallon

All sentences were presented using a paper-and-pencil format with the same phrasing as Stimulus Sentence Set 1 of Experiment 1.

### Procedure

Session 1. Subjects were run in small groups of up to 12 people during each of the experimental sessions. The written instructions told the subjects to circle "the one choice (of the four) that conveys the equivalent information as the original statement." Each of the four contexts were given, along with the set of multiple choice alternatives corresponding to that context. Subjects were told that each of the sentences dealing with the same context were somewhat different from the others. Subjects were not under any time constraints, and spent between 15 and 30 minutes performing the task in Session 1.

Session 2. The subjects returned one week later. The two experimental groups saw the videotape lesson and then did the problems. The monolingual control group was given only the sentence task.

The 35 minute videotaped lesson had two specific goals. First, it taught subjects to use a procedure for removing negations by paraphrasing negative portions into the affirmative by proceeding from right to left. The second goal was to eliminate the strategy of translating "not all" or "it's not true that all" to "some" by drawing diagrams portraying the meaning of the sentence.

The lesson began with the following sample item:

"It's not true that not all tires on sale at the Acme Auto Supply are whitewall tires."

a) All tires on sale at the Acme Auto Supply are whitewall tires.

b) Some tires on sale at the Acme Auto Supply are whitewall tires.

c) All tires on sale at the Acme Auto Supply are blackwall tires.

d) Some tires on sale at the Acme Auto Supply are blackwall tires.

Using this sample item as a concrete example, the general structure of the sentences was depicted as follows:

|                        |             |                     |
|------------------------|-------------|---------------------|
|                        | [ some ]    |                     |
| [ It's not true that ] | [ all ]     | [ group of things ] |
| optional               | [ not all ] |                     |

|           |                    |
|-----------|--------------------|
|           | [ Category A ]     |
| belong to | [ Category not A ] |

These parts were delineated in detail for the sample item. Following the explanation of the general structure of the sentences, a four-step procedure for processing the sentences into affirmative statements was outlined:

1. Make sure you know what the two categories are; that is, category "A" and category "not A" (for "not A," the category complement was always stated orally).  
example - whitewall tires and not whitewall tires, or blackwall tires
2. Start from the end of the problem and work toward the beginning until a negation is encountered.
3. Get rid of the negation by paraphrasing from the point where the negation occurs to the end of the problem.

4. Repeat steps 2 and 3 until the problem statement is paraphrased into a totally affirmative statement.

This procedure was demonstrated in detail for four of the eight Sentence Types used in the study. At the end of the lesson, the subjects were given two sample sentences to process on their own which were of a different type from those covered in the lesson (thus, subjects either saw examples of, or worked out, six different Sentence Types during the course of the video lesson). The correct solutions for the two sample sentences were then reviewed on the videotape.

The critical step in the procedure was step #3, since subjects need to be taught how to paraphrase negative portions of sentences according to their underlying meaning and not their surface structure. Therefore, rough diagrams were used in conjunction with the sample items. For example, consider the following portion of the sample sentence: "not all tires on sale at the Acme Auto Supply are whitewall tires." The lesson said:

"There may be a lot of tires on sale at the Acme Auto Supply, but for the sake of argument, suppose that there are only a total of five tires on sale. Here are these five tires (instructor draws five marks as shown below).

| | | | |

To say that not all of these five tires are whitewall tires means that perhaps one (points to one of the marks), or perhaps three (points to three marks) or perhaps even all five (points to all five marks) tires are blackwall tires, since we know that all five can't be whitewalls. So to say that not all tires are whitewall tires must mean that some tires, whether it's one, three or perhaps all of them you don't really know, but at least some tires have to be blackwalls."

At this point, the subjects were cautioned that "not all tires are whitewalls" does not mean exactly the same thing as "some tires are whitewalls," since all the tires could be blackwalls.

Immediately following the videotape the experimental subjects were presented with the 32 item-task. They were asked to use the method they learned during the videotaped lesson to process the sentences, and to indicate in as much detail as possible in the booklet the transformations that they used to paraphrase the sentences into affirmative statements.

Session 3. One week later, the subjects returned to work the 32 items a third time. The subjects who had seen the videotape were again asked to show in detail the transformations they used to paraphrase the sentences.

Session 4. Six months after the first session, 11 of the bilingual and 14 of the monolingual experimental subjects returned to take the test a fourth time, in order to study the long-term retention of the approach taught in the videotaped lesson. No members of the control group returned for this fourth session.

## Results

### Preliminary Analysis of Session 1

As in Experiment 1, it was desirable to combine the two Continuous Contexts, 1 and 2, and the two Discrete Contexts, 3 and 4, in order to simplify the interpretation of later analyses. A comparison of the two Continuous Contexts revealed no difference in percent correct: the means were .46 for Context 1, and .48 for Context 2. However, there was a difference in percent correct between the two Discrete Contexts,  $F(1,54) = 4.06$ ,  $p=.0488$ : the mean for Context 3 was .58, and for Context 4 was .53. Given the small

size of the difference between these means and the potential simplification of the analyses, the two Discrete Contexts were combined despite this marginally significant difference.

#### Session 1 - Comparison to Experiment 1

Group. In contrast to Experiment 1, a Group (BE, ME, MC) x Sentence (8 types) x Context (Discrete, Continuous) ANOVA on percent correct scores in Session 1 revealed a main effect of Group,  $F(2,54) = 4.62, p=.0141$ . There was no difference between the two monolingual groups, ME and MC,  $F(1,38) = .00, p=.9831$ . However, the performance of the bilingual group, 39%, was significantly lower than the performance of the combined M groups, 57%,  $F(1,55) = 8.73, p(\text{Bonferroni}, k=2) <.01$ . This difference may have resulted from the generally lower ability of the students in Experiment 2 (as indicated by lower mean GPA) accentuating the trend toward a lower performance by the bilingual group.

Sentence Type. As in Experiment 1, there was a main effect of Sentence Type,  $F(7,378) = 4.84, p<.0001$ . The pattern of results was similar to that observed in Experiment 1, although the means were somewhat lower (Compare Figures 1 and 4). The mean percent correct on Sentence Types with one negation was 54%, followed by two negative Sentence Types, 52%, and the three negative Sentence Type, 41%. A Group x Negations x Context ANOVA revealed a main effect of number of Negations,  $F(2,108)=4.01, p=.0210$ . However, none of the pairwise comparisons were significant at the  $p=.0167$  needed for significance with 3 comparisons.

-----  
INSERT FIGURE 4 ABOUT HERE  
-----

Context The main effect of the Context observed in Experiment 1 was replicated in Experiment 2,  $F(1,54)=20.97$ ,  $p<.0001$ . The mean performance on Discrete items, .56, was significantly better than performance on Continuous items, .47. There was again an interaction of Sentence Type and Context,  $F(7,378)=6.53$ ,  $p<.0001$ : performance was better in the Discrete Context for all Sentence Types but 6. The difference between the Discrete and Continuous Contexts was significant (with Bonferroni test,  $k=8$ ) for Sentence Types 1-  $t(56)=4.774$ ,  $p<.01$ , 3-  $t(56)=3.459$ ,  $p<.01$ , and 8-  $t(56)=5.981$ ,  $p<.01$ .

There was also an interaction between Context and number of Negations,  $F(2,108)=11.93$ ,  $p<.0001$ . As in Experiment 1, performance was significantly better (Bonferroni,  $k=3$ ) for single negation Sentence Types in the Discrete Context,  $t(56)=4.45$ ,  $p<.0001$ . In addition, performance was better for two negation Sentence Types in the Discrete Context,  $t(56)=4.14$ ,  $p<.0001$ . There was no difference between Contexts for the three negation Sentence Type.

#### Session 2 - Immediate Effects of the Lesson

Time There was a dramatic increase in mean performance in Session 2: the overall mean for the three groups was 80%, versus 51% for Session 1 (See Table 3). This difference between sessions was highly significant,  $F(1,54)=81.74$ ,  $p<.0001$ . However, the difference in performance between Sessions 1 and 2 was not the same for all groups,  $F(2,54)=12.35$ ,  $p<.0001$ .

Group The performances of both the BE and the ME groups were significantly higher in Session 2 than in Session 1. The mean performance of the BE group went from 39% to 79%,  $t(16) = 5.208$ ,  $p(\text{Bonferroni}, k=3) <.0001$ , while the mean performance of the ME group went from 57% to 97%,  $t(19) = 7.958$ ,  $p(\text{Bonferroni}, k=3) <.0001$ . A comparison of the ME and BE groups in a separate ANOVA, showed the ME group scored significantly higher than the BE

group,  $F(1,35)=8.45$ ,  $p=.0063$ . However, the lack of a Time x Group interaction when BE and ME are compared indicates the groups made similar gains in performance. In contrast, the monolingual control group improved only from 56 % to 64%, a difference that was not significant.

-----

INSERT Table 3 ABOUT HERE

-----

There was a main effect of Group in Session 2,  $F(2,48) = 27.68$ ,  $p<.0001$ : the performance of the ME group, mean 97%, was better than the performance of both the BE group, mean 79%  $t(35) = 2.91$ ,  $p(\text{Bonferroni}, k=3)=.02$ , and the MC group, mean 64%,  $t(38) = 6.089$ ,  $p(\text{Bonferroni}, k=3) <.0001$ . However, the 15% difference (79% versus 64%) between the performances of the BE and MC groups was not significant (See Table 3).

The very high performance of the ME group in Session 2 indicates that most of these subjects had attained ceiling level. In fact, an examination of the individual scores reveals that no subject scored lower than 28 (of 32) correct. However, the data for the BE group is considerably more varied: eleven of the subjects scored at least 29 correct, three scored between 21 and 23, and three scored 11 or lower. Any subject who scored lower than 85% correct presumably did not learn to implement the parsing method correctly. Therefore, a fair test of retention would not include those subjects who did not display competency with the parsing method. Accordingly, the six BE subjects who scored 23 or lower were dropped from further analyses. The results for these subjects will be discussed separately at the end of this section. Once these subjects are dropped, the mean for the BE group becomes .96 and there is no longer a significant difference in performance between the BE and ME groups in Session 2.

The performance of the MC group also improved from Session 1 to Session 2, 57% versus 64%. However, this increase was not significant: the performance of both the BE group and the ME group, was significantly better than that of the MC group,  $t(29) = 4.4125$ ,  $p(\text{Bonferroni}, k=3) = .0003$ , and  $t(38) = 6.089$ ,  $p(\text{Bonferroni}, k=3) < .0001$ .

Sentence Type and Context Given that the two groups taught the parsing method, BE and the ME, were nearly at ceiling, it is not surprising that there was no longer an effect of either Sentence Type or Context in session 2.

### Session 3 - Retention One Week Following the Lesson

Time. One week after the parsing lesson, there was no difference in overall performance from the immediate posttest. In fact, rather than any decrement in performance as one might expect, performance actually was somewhat better in Session 3: the mean for all subjects in session 3 was 86% versus 84% in Session 2.

Group. The Group effect seen in Session 2 was also present in Session 3,  $F(2,48) = 22.74$ ,  $p < .0001$ . The mean performance of the BE group was 98%, of the ME group was 98%, and of the MC group was 67%. Again, both the BE group and the ME group performed better than the MC group,  $t(29) = 4.054$ ,  $p(\text{Bonferroni}, k=3) = .0009$ , and  $t(38) = 5.435$ ,  $p(\text{Bonferroni}, k=3) < .0001$ , but were not different from each other. For none of the groups was there any difference in performance between Sessions 2 and 3, although performance did improve very slightly for each group.

In addition to a significant increase in overall level of accuracy from Session 1 to Session 2,  $F(1,48) = 162.34$ ,  $p < .0001$ , there was a significant increase in accuracy for each group from Session 1 to Session 3. Group BE's

performance went from 40% to 98%,  $t(10) = 11.216$ ,  $p(\text{Bonferroni}, k=3) < .0001$ , Group ME rose from 57% to 98%,  $t(19) = 8.265$ ,  $p(\text{Bonferroni}, k=3) < .0001$ , while Group MC's performance rose from 56% to 67%,  $t(18) = 2.673$ ,  $p(\text{Bonferroni}, k=3) = .04$ . This gain in performance was not the same for each group,  $F(2,48) = 7.47$ ,  $p < .0001$ ; the experimental groups gained considerably more than the control group. The eleven point gain in performance by the control group is probably close to the limit that performance would improve through only repeated administration of the task. Both the BE group and the ME group remained superior to the MC group in this third session,  $t(29) = 4.054$ ,  $p(\text{Bonferroni}, k=3) = .0009$  and  $t(37) = 4.656$ ,  $p(\text{Bonferroni}, k=3) < .0001$ . There was no difference between the BE and the ME groups.

Sentence and Context Effects As in session 2, there were neither Sentence Type effects nor Context effects. Once the subjects had learned the parsing method, they were able to apply it equally well for all items.

#### Session 4 - Long-term Retention

Time. After a six month period, 8 members of the BE group and 14 members of the ME group were still at the university and returned for an assessment of long term retention. Averaging over group, there was a significant decrement over time from Sessions 3 to 4,  $F(1,20) = 14.32$ ,  $p = .0012$ . For the subjects who returned, performance dropped from 98% in Session 3 to 84% in Session 4. However, performance in Session 4 was still considerably better than performance in Session 1: 84% versus 54%,  $F(1,20) = 18.19$ ,  $p = .0004$ . Thus, although retention of the parsing method was not 100%, performance in the long term was clearly improved over the original performance.

Group. There was a difference between the two groups in how well they retained the parsing method from Session 3, i.e., there was a Group x Time interaction,  $F(1,20)=9.10$ ,  $p=.0068$ . The decrement in performance for the BE group was 31 points (98% versus 66%), which was considerably larger than the 4 point (98% versus 95%) decrement for the ME group. In fact, with the exception of one ME subject who scored 19 of 32 (59%) correct in Session 4, the ME group showed no decrement.

In contrast, only 3 of the 8 members of the BE group scored above 85% in Session 4: the other subjects scored 78%, 66%, 44%, and 25%. The decrement in performance from Sessions 3 to 4 was significant for the BE group,  $F(1,7) = 7.41$ ,  $p=.0297$ . There was an overall improvement in performance from Session 1 to 4 for the BE group, 45% to 66%. However, this difference was not significant, due to the size and variability of the group.

#### Strategies Related to Poor Performance

The videotaped lesson on the parsing strategy was clearly effective for the majority of the subjects. However, a sizeable subset of the BE group, 6 of 17, did not display proficiency with the task immediately following the lesson. One of the 6 subjects subsequently scored 30 of 32 correct on the posttest one week later, but none of the others made such a dramatic improvement. Clearly, these subjects used alternative strategies that did not result in correct performance. Fortunately, the subjects were told to use the method taught in the videotape lesson and show their work in the blank space provided. Hence, for the most part, it is possible to determine what strategies the subjects used to obtain their answers.

The majority of the errors made in Session 2 (95 of 110) were made by switching the quantifier "all" to "some" or "some" to "all" when preceded by a

falsification or a subject negation, without concurrently changing the category to the opposite category. For example, "Not all clerks ... are male" would become "Some clerks ... are male." Quantifier switching could have occurred in sentences that contained "not all", "It's not true that all", and "It's not true that some." The phrase "not all" occurred in four sentence types, while "It's not true that all" and "It's not true that some" each occurred in two sentence types (i.e., the frequencies of occurrence were in the ratio 4:2:2). The actual ratio of errors was 59:17:19, suggesting that subjects seemed slightly more predisposed to make an error interpreting "not all" than in interpreting either of the "It's not true" cases. Seven of the remaining errors were made by changing the category, but not the quantifier when presented with "It's not true that all...". The other eight errors were either simply mistakes in answering or were not classifiable.

The pattern of errors for the five subjects in the BE group that continued to perform below 85% in Session 3 was similar to the pattern of errors made in Session 2. Of 82 errors, 41 were due to equating "not all" with "some", 13 were due to equating "It's not true that all" with "It is true that some", and 22 were due to equating "It's not true that some" with "It is true that all." There were 4 category change errors in response to "It's not true that all..", and 2 unclassifiable errors.

What is of considerable interest about these error patterns is that they are quite similar to the error patterns of subjects who performed well in the immediate and one-week posttests, but below 85% 6 months later. Six subjects, five from the BE group and one from the ME group, are included in these figures. Of the 95 errors, 44 were "not all" to "some" errors, 24 were "It's not true that all.." goes to "some" errors, and 24 were "It's not true that some.." goes to "all" errors. Three errors were not classifiable. Both

experiments indicate that subjects initially approach the task with a collection of natural language strategies. Most subjects abandon these strategies when presented with the videotaped lesson and apply the parsing strategy correctly the majority of the time. However, failure to recall the salient features of the lesson leads to use of the former natural language strategies.

### Discussion

The findings of Session 1 in Experiment 2 replicated the major effects of Experiment 1: 1) One- and two-negation sentences were significantly easier than three-negation sentences, but one-negation sentences were not easier than two-negation sentences, 2) Discrete Contexts were significantly easier than Continuous Contexts, and 3) there was no Group by Negation interaction. However, there was one important difference. In Session 1 of Experiment 2, the accuracy of the bilingual group (39%) was considerably poorer than that of both the two monolingual groups (57%), as well as comparable bilingual (54%) and monolingual (60%) groups of Experiment 1. This poor performance is likely due to the fact that the bilinguals of Experiment 2 were academically underprepared as measured by both GPA and SAT (See Table 2). The low SAT-Verbal scores can reflect not only limited English proficiency, but also insufficient knowledge of English for academic purposes as represented by the linguistic demands of this test.

The intervention strategy implemented in Session 2 resulted in dramatic, long lasting improvements in performance; except for a subset of the Hispanic group who did not exhibit beneficial effects, the lesson gave rise to ceiling level performance. Given subjects' initial difficulties with this task, and

the relatively short duration of the videotaped lesson, the prolonged positive effect on performance is quite encouraging.

In designing the intervention strategy, we attempted to take into account a number of findings suggested by cognitive research from various domains. These included: 1) In a similar experiment, the only subjects who consistently obtained the correct answer used a right-to-left parsing strategy (Mestre, 1986), 2) Subjects learn a procedure best by practicing it (Anzai and Simon, 1979), and 3) Misconceptions found in domains such as physics (Champagne, Klopfer and Gunstone, 1982), algebra (Clement, 1982; Mestre, Gerace and Lochhead, 1982) and statistics (Pollatsek, Lima, and Well, 1981; Tversky and Kahneman, 1977), are not only deeply lodged, but also interfere with cognitive functioning in that domain. Thus, the focus of the intervention strategy was to demonstrate a procedural right-to-left parsing method for decomposing and understanding the different portions of the sentences. This method was supplemented by discussions of common misconceptions (such as simply substituting "not all" for "some") in an attempt to help students recognize, and dislodge these misconceptions. The false logic giving rise to the misconceptions was illustrated as incorrect through the use of concrete examples and diagrams. Sample problems were also provided during the course of the lesson so that subjects could practice and assimilate the parsing procedure.

The intervention strategy proved to be much more effective in both the long, and short run for the monolingual group than for the bilingual group. Whereas the performance of all monolingual subjects was nearly perfect immediately following the videotaped lesson, only two-thirds of the bilingual subjects were able to match this performance level. In addition, thirteen out of the fourteen monolinguals who returned for the six-month posttest exhibited

near perfect performance, whereas only three out of the eight returning bilinguals were able to reach ceiling-level performance. This disparity in performance may have two sources.

The first source concerns the Hispanic group's relative disadvantage in English proficiency as indicated by the substantially lower verbal performance in verbal SAT scores. This language deficiency would likely be a handicap in an intervention strategy that not only was delivered verbally via a videotaped lesson, but that also relied heavily upon subjects' ability to distinguish several subtleties in meaning among some rather complex sentences. This view is consistent with a theoretical framework proposed by Cummins (1979) on the cognitive functioning of bilinguals. Cummins proposes a "linguistic threshold hypothesis" which posits that "there may be a threshold level of linguistic competence which bilingual children must attain both in order to avoid cognitive deficits and to allow the potentially beneficial aspects of becoming bilingual to influence their cognitive growth" (p. 229). It appears that the linguistic proficiency level of the bilingual Hispanics of Experiment 2 was sufficiently "below threshold" (in the Cummins sense) to impede their performance, given the linguistic complexity of the tasks in experiment 2.

Second, it is clear that subjects possess a number of misconceptions concerning the interpretation of certain negated phrases which interfere with performance. These misconceptions are based on strategies that are appropriate for comprehending natural discourse but which are inappropriate in formal logic. The analysis of the incorrect strategies used by subjects in the sessions following the videotaped lesson indicates that these misconceptions are deeply seated and not easily dislodged for a substantial number of the Hispanic subjects. The most convincing evidence for this view comes from the similarity of the error patterns exhibited in Sessions 2 and 4.

The error patterns of those subjects who exhibited a ceiling-level performance in the Session 2 immediately following the lesson, but who performed poorly in Session 4 six months later, were nearly identical to the error patterns of those subjects who performed poorly immediately following the lesson. Thus, even among those Hispanic subjects who were successful in abandoning their misconceptions in Session 2 immediately following the lesson, there appeared to be a strong predisposition toward reverting to these misconceptions after a prolonged period of time. These findings are consistent with research findings from domains such as physics (Champagne, Klopfer and Gunstone, 1982) and algebra (Clement, 1982) which indicate that misconceptions are deep seated and not easily supplanted.

There is one distinction between misconceptions in content areas, such as physics and algebra, and the type of semantic misconceptions that we have been discussing (e.g. not all clerks are male = some clerks are male). Misconceptions in a content area may impact only on learning more about that specific content area. However, semantic misconceptions can have an impact on learning subject matter across all content areas.

#### GENERAL DISCUSSION

The findings of Experiment 1, in accordance with previously discussed research, indicate that negations are hard to comprehend. This may be true for several reasons. First, the presence of a negation places strain on comprehension mechanisms, as indicated by a consistent increment in processing time as the number of negatives is increased. Second, subjects appear to use a number of inappropriate strategies to cope with this strain: these strategies are often misapplied rules of logic or mathematics (such as cancelling two negations) or a misuse of the rules governing comprehension of

natural discourse (such as substituting "some" for "not all"). These misconceptions that subjects bring to comprehension tasks interfere with performance. The vast majority of the errors observed on this task were caused by a small number of error types common across a wide range of subjects, which appeared either singly, or in combinations of two or more. This pattern of faulty logic due to misconceptions is paralleled in other domains, such as physics (Champagne et al., 1982; McCloskey, 1983). This research indicates that novices possess beliefs about the physical world which are contrary to the tenets of classical mechanics and interfere with learning. These misconceptions are often the result of attempts to explain physical phenomena based on incomplete or inappropriately applied information.

In comprehending the text that appears in Experiment 1, subjects may naturally assume that the rules for comprehending natural discourse are applicable. Thus, a statement such as "not all clerks are male" would be interpreted to mean "some clerks are male" since that would be the intended meaning in conversation. Normally, the speaker would state "some clerks are female," or "all clerks are female," if either of these were the intended meaning. Similarly, young students are taught that "two negatives make a positive" and are admonished never to say "I don't want nothing," since formally it means "I want something." Therefore, it is not surprising that the subjects cancelled negatives in pairs when the opportunity arose.

The intervention strategy of Experiment 2 was quite effective both in the short- and long-term, a result that is quite encouraging. However, the intervention strategy proved to be much more effective for the monolingual than for the Hispanic subjects: several Hispanic subjects showed little indication that they had initially learned the parsing procedure, while several others appeared to have forgotten the procedure after six months. One

explanation for this disparity in performance is the relative difference in the level of language proficiency between the two groups. The Hispanic subjects scored substantially below the monolingual group in Verbal SAT, implying a lower proficiency in English for academic purposes. This lower level of proficiency may have been a factor in subjects grasping only part of the parsing method, "forgetting," for example, the caution about the illegality in substituting "some" for "not all." In addition, other subjects may have initially been able to perform using the rules, but not understood them to the same degree that the monolinguals did. Therefore, the rules were not retained after six months. It is interesting to note that the Hispanics who initially performed well immediately following the lesson, but performed poorly six months later exhibited the same misconceptions as those Hispanic subjects who performed poorly immediately following the lesson. This implies, not surprisingly, that the ability to dislodge misconceptions based on subtleties in language meaning depends on the individual's language proficiency level.

These experiments have several educational implications, one of which concerns the effect that negations have on the performance of the typical undergraduate. There is a prevalence of single and double negations in both the written and oral communication intended for college students and the general public. Negations are not generally uncommon in newspaper and magazine articles, textbooks, newscasts, lectures, etc. These negations appear to blend in with the communication so that they often go unnoticed, as illustrated by the single negation in this sentence and the double negation in the previous sentence. Our research has made clear that negatively phrased sentences are significantly harder to comprehend than affirmatively phrased sentences with similar meanings. When negatives appear frequently in course

lectures and tests, it is likely that many students, particularly low-verbal students, have a harder time understanding the message than if it were phrased in the affirmative. It is important that as educators we are cognizant of this.

Perhaps the most significant educational implication of this research concerns instructional design. The design of the intervention strategy of Experiment 2 utilized several important factors which may be generalized to other domains. We feel that the key to designing an effective intervention strategy is to perform a cognitive analysis of the task in order to identify correct as well as erroneous strategies used by the population of interest. The correct strategy should be illustrated in a series of easily comprehensible steps. Equally important is illustrating common misconceptions that could interfere with performance, since it is only by confronting misconceptions that one is able to dislodge them. Finally, the acquisition of any procedure is not complete without practice in performing the task and receiving feedback.

#### Acknowledgements

The research reported in this article was supported by National Institute of Education grant #NIE-83-0072. The contents herein do not necessarily reflect the position, policy or endorsement of NIE. We would like to thank Prof. Charles Clifton for many helpful discussions, Emily Marion for assistance in running Experiments 1 and 2, and Ron Marion for technical assistance in Experiment 1.

## References

- Akiyama, M.M. (In Press) Denials in young children from a cross-linguistic perspective. Child Development.
- Akiyama, M.M. (1994) Our language acquisition strategies universal? Developmental Psychology, 20, 219-228.
- Anzai, Y. and Simon, H.A. (1979) The theory of learning by doing. Psychological Review, 86, 124-140.
- Arroyo, F.V. (1982) Negatives in context. Journal of Verbal Learning and Verbal Behavior, 21, 118-126.
- Braine, M.D.S. (1978) On the relation between the natural logic of reasoning and standard logic. Psychological Review, 85, 1-21.
- Carpenter, P.A. and Just, M.A. (1975) Sentence comprehension: a psycholinguistic processing model of verification. Psychological Review, 82, 45-73.
- Champagne, A.B., Klopfer, L.E., and Gunstone, R.F. (1982) Cognitive research and the design of science instruction. Educational Psychologist, 17, 31-53.
- Chapman, L.J. and Chapman, J.P. (1959) Atmosphere effect reexamined. Journal of Experimental Psychology, 58, 220-226.
- Chi, M.T.H., Feltovich, P.J., and Glaser, R. (1981) Categorization and representation of physics problems by experts and novices. Cognitive Science, 5, 121-152.
- Clark, H.H. and Card, S.K. (1969) The role of semantics in remembering comparative sentences. Journal of Experimental Psychology, 82, 545-553.

- Clement, J. (1982) Algebra word problem solutions: Thought processes underlying a common misconception. Journal for Research in Mathematics Education, 13, 16-30.
- Cummins, J. (1979) Linguistic interdependence and the educational development of bilingual children. Review of Educational Research, 49, 222-251.
- Gough, P.B. (1965) Grammatical transformations and speed of understanding. Journal of Verbal Learning and Verbal Behavior, 4, 107-111.
- Henle, M. (1962) On the relation between logic and thinking. Psychological Review, 69, 366-378.
- Hunt, E. (1978) Mechanics of verbal ability. Psychological Review, 85, 109-130.
- Johnson-Laird, P.N. (1983) Mental Models: Toward a Cognitive Science of Language, Inference, and Consciousness. Cambridge, MA: Harvard University Press.
- Johnson-Laird, P.N. (1970) Linguistic complexity and insight into a deductive problem. In G.B. Flores d'Arcais & W.J.M. Levelt (Eds.) Advances in Psycholinguistics. Amsterdam: North Holland.
- Johnson-Laird, P.N. and Steedman, M. (1978) The psychology of syllogisms. Cognitive Psychology, 10, 64-99.
- Just, M.A. and Carpenter, P.A. (1971) Comprehension of negation with quantification. Journal of Verbal Learning and Verbal Behavior, 10, 244-253.
- Legrenzi, P. (1970) Relations between language and reasoning about deductive rules. In G.B. Flores d'Arcais and J.J.M. Levent (Eds.). Advances in Psycholinguistics. Amsterdam: North Holland.
- McCloskey, M. (1983) Intuitive physics. Scientific American, 248.

- Mehler, J. (1963) Some effects of grammatical transformations on the recall of English sentences. Journal of Verbal Learning and Verbal Behavior, 2, 346-351.
- Mestre, J.P. (1981) Predicting academic achievement among bilingual Hispanic college technical students. Educational and Psychological Measurement, 41, 1255-1264.
- Mestre, J.P. (1982) "Just how important are language skills in mathematical problem solving?" In S. Wagner (Ed.) Proceedings of the Fourth Annual Meeting of the North American Chapter of the Psychology of Mathematics Education. 126-132. Athens, Dept. of Mathematics Ed., University of Georgia.
- Mestre, J.P. (1986) The Latino science and engineering student: Some recent research findings. In M. Olivas (Ed.), Latino College Students. New York: Columbia University Press.
- Mestre, J.P., Gerace, W., and Lochhead, J. (1982) The interdependence of language and translational math skills among bilingual Hispanic engineering students. Journal of Research in Science Teaching, 19, 399-410.
- Miller, G.A. (1962) Some psychological studies of grammar. American Psychologist, 17, 748-762.
- Miller, G.A. and McKean, K.A. (1964) Chronometric study of some relations between sentences. Quarterly Journal of Experimental Psychology, 16, 297-308.
- Pollatsek, A., Lima, S.D., and Well, A. (1981) Concept or computation: Students' understanding of the mean. Educational Studies in Mathematics, 12, 191-204.

- Schoenfeld, A.H. and Herrmann, D.J. (1982) Problem perception and knowledge structure in expert and novice mathematical problem solvers. Journal of Experimental Psychology: Learning, Memory, and Cognition, 8, 484-494.
- Sherman, M.A. (1973) Bound to be easier? The negative prefix and sentence comprehension. Journal of Verbal Learning and Verbal Behavior, 12, 76-84.
- Sherman, M.A. (1976) Adjectival negation and the comprehension of multiply-negated sentences. Journal of Verbal Learning and Verbal Behavior, 15, 143-157.
- Slobin, D.I. (1965) Grammatical transformations and speed of understanding. Journal of Verbal Learning and Verbal Behavior, 5, 219-227.
- Trabasso, T., Rollin, H., and Shaughnessey, E. (1971) Storage and verification shapes in processing concepts. Cognitive Psychology, 2, 231-289.
- Tversky, A. and Kahneman, D. (1977) Judgment under uncertainty: Heuristics and biases. In P.N. Johnson-Laird and P.C. Wason (Eds) Thinking: Readings in Cognitive Science, Cambridge: Cambridge University Press, 326-337.
- Vazquez, C.A. (1981) Sentence processing: Evidence against the serial, independent stage assumption. Journal of Psycholinguistic Research, 10, 363-374.
- Wason, P.C. (1959) The processing of positive and negative information. Quantitative Journal of Experimental Psychology, 11, 92-107.
- Wason, P.C. (1961) Response to affirmative and negative binary statements. Journal of Psychology, 52, 133-142.
- Wason, P.C. (1965) The contexts of plausible denials. Journal of Verbal Learning and Verbal Behavior, 4, 7-11.

Wason, P.C. and Johnson-Laird, P.N. (1972) Psychology of Reasoning: Structure and Content. Cambridge, MA: Harvard University Press.

Woodworth, R.S. and Sells, S.B. (1935) On atmospheric effect in formal syllogistic reasoning. Journal of Experimental Psychology, 18, 451-460.

Table 1: Eight Sentence Types and Four Multiple Choice Alternatives

- 1) Not all of the clerks working at the Fitzgerald Company are male.  
(d)
- 2) Not all of the clerks working at the Fitzgerald Company are not male. (b)
- 3) It's not true that all of the clerks working at the Fitzgerald Company are male. (d)
- 4) It's not true that all of the clerks working at the Fitzgerald Company are not male. (b)
- 5) It's not true that not all of the clerks working at the Fitzgerald Company are male. (a)
- 6) It's not true that not all of the clerks working at the Fitzgerald Company are not male. (c)
- 7) It's not true that some of the clerks working at the Fitzgerald Company are male. (c)
- 8) It's not true that some of the clerks working at the Fitzgerald Company are not male. (a)

Four Multiple Choice Alternatives:

- a) All of the clerks working at the Fitzgerald Company are male.
- b) Some of the clerks working at the Fitzgerald Company are male.
- c) All of the clerks working at the Fitzgerald Company are female.
- d) Some of the clerks working at the Fitzgerald Company are female.

Table 2: Mean SAT Scores and GPA's  
(Standard Deviations in Parentheses)

| Group                       | SAT Quantitative | SAT Verbal | GPA        |
|-----------------------------|------------------|------------|------------|
| Bilingual                   | 426 (94)         | 333 (83)   | 2.30 (.40) |
| Monolingual<br>Experimental | 620 (72)         | 502 (96)   | 2.79 (.67) |
| Monolingual<br>Control      | 634 (74)         | 520 (95)   | 2.88 (.51) |

Table 3: Mean Group Performances Across Sessions

| Group                    | Session |     |     |     |
|--------------------------|---------|-----|-----|-----|
|                          | 1       | 2   | 3   | 4   |
| Bilingual                |         |     |     |     |
| (n=17)                   | .39     | .79 | --  | --  |
| (n=11) *                 | .40     | .96 | .98 | --  |
| (n=8)+                   | .45     | .96 | .98 | .66 |
| Monolingual Experimental |         |     |     |     |
| (n=20)                   | .57     | .97 | .98 | --  |
| (n=14)+                  | .60     | .98 | .98 | .95 |
| Monolingual Control      |         |     |     |     |
| (n=20)                   | .56     | .64 | .67 | --  |

---

\* Subjects who scored above .85 in session 2

+ Subjects who returned 6 months later for session 1V

Figure 1

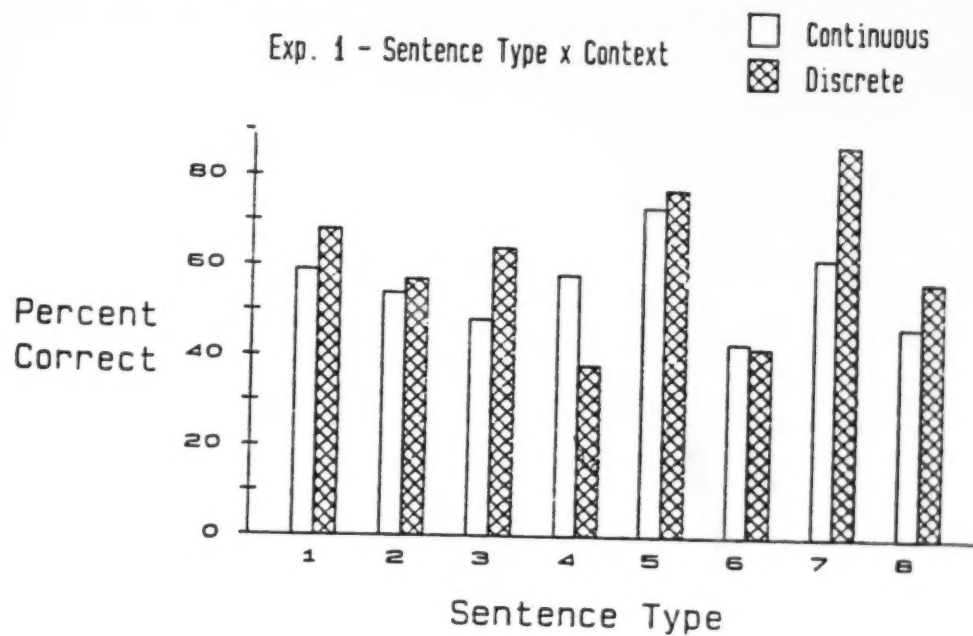


Figure 2

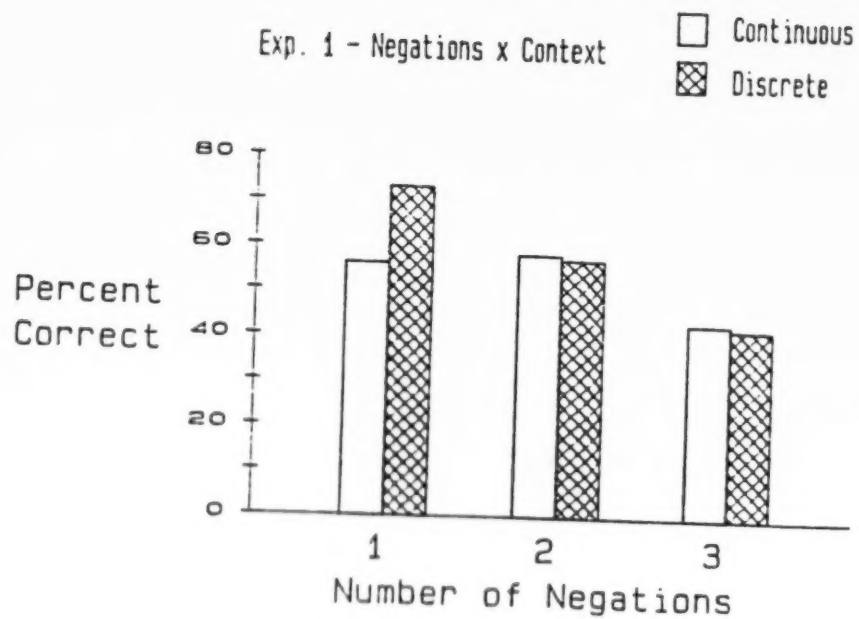


Figure 3

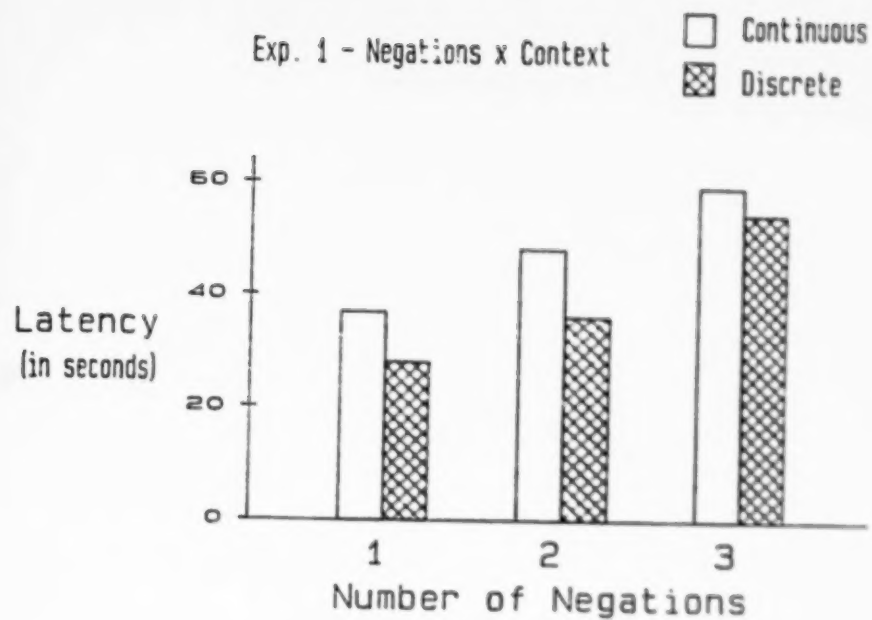


Figure 4

Exp. 2 - Sentence Type x Context

Continuous  
Discrete

